

<https://helda.helsinki.fi>

---

## Phylogenetic comparative method for geographical radiation

Ramiadantsoa, Tanjona

2017-04

---

Ramiadantsoa , T , Siren , J & Hanski , I 2017 , ' Phylogenetic comparative method for geographical radiation ' , Annales Zoologici Fennici , vol. 54 , no. 1-4 , pp. 237-257 . <https://doi.org/10.5735/086.054.>

---

<http://hdl.handle.net/10138/236230>

<https://doi.org/10.5735/086.054.0121>

---

publishedVersion

---

*Downloaded from Helda, University of Helsinki institutional repository.*

*This is an electronic reprint of the original article.*

*This reprint may differ from the original in pagination and typographic detail.*

*Please cite the original version.*

**Ilkka Hanski: The legacy of a multifaceted ecologist**

## Phylogenetic comparative method for geographical radiation

Tanjona Ramiadantsoa<sup>1,2,3,\*</sup>, Jukka Sirén<sup>1</sup> & Ilkka Hanski<sup>1,†</sup>

<sup>1</sup> *Metapopulation Research Centre, Department of Biosciences, P.O. Box 65, FI-00014 University of Helsinki, Finland*

<sup>2</sup> *Department of Ecology, Evolution, and Behavior, University of Minnesota, 1987 Upper Buford Cir Saint Paul, MN 55108-6097, USA*

<sup>3</sup> *Department of Zoology, University of Wisconsin-Madison, 430 Lincoln Dr, Madison, WI 53706, USA (\*corresponding author's e-mail: ramiadantsoa@wisc.edu)*

*Received 25 Oct. 2016, final version received 16 Mar. 2017, accepted 17 Mar. 2017*

Ramiadantsoa, T., Sirén, J. & Hanski, I. 2017: Phylogenetic comparative method for geographical radiation. — *Ann. Zool. Fennici* 54: 237–257.

Phylogeny can provide information about the processes that have shaped extant diversity. Here, we complement existing comparative phylogenetic methods by developing a model that couples diversity-dependent diversification rate and range dynamics. Unlike many models, we used Approximate Bayesian Computation to fit the model to the data. We validated the inference by estimating known parameter values from simulated data, and found that within-region speciation and extinction rates cannot be simultaneously estimated most likely due to correlations among parameter values. Since the model can estimate a diversification rate, we applied the model to a monophyletic lineage of 74 species of dung beetles (Canthonini: *Nanos* and *Apotolamprus*) endemic to Madagascar. The estimated diversification rate is clearly higher in northern than in eastern or western Madagascar. The current species richness is highest in North where complex topography and a mixture of biomes likely favour ecological diversification. The approach we have developed here is a step towards examining weaknesses and strengths of phylogenetic comparative methods in an explicit spatial context. Further development and testing of the model is needed before its routine application to empirical data.

## Introduction

Speciation and ecological diversification lead to an increase in the number of species in a clade until opportunities for establishment of further species with sufficiently dissimilar specializations are exhausted. In the classic examples of such evolutionary radiations, including the Galapagos finches (Grant 1999), the Hawaiian honeycreepers (Freed *et al.* 1987, Futuyma 1998,

Lovette *et al.* 2002), and the East African cichlid fishes (Verheyen 2003, Seehausen 2006, Day 2008), radiation has occurred within a relatively small area, and the focus of the research has been in the evolution of divergent morphological traits of species occupying distinct ecological niches. Other radiations have taken place in systems of discrete areas, such as islands and caves, which have provided an opportunity to ask questions about the similarity of independent radiations

within separate areas (Mahler *et al.* 2013). In such systems, but also within large continuous areas, most species have geographical ranges that are smaller than the total area in question. A new species may evolve within a small area, after which it may expand its geographical range, while the global extinction of a species is likely to be preceded by a shrinkage of its range. Range dynamics are occasionally fast, especially when a species colonizes a new region without close competitors and predators [for an informative example on Malagasy dung beetles *see* Hanski *et al.* (2008)], but perhaps more frequently ranges expand and shrink slowly, often because interspecific interactions constrain the spatial spreading of species (Waters *et al.* 2013). In that latter case, speciation and extinction rates on the one hand, and the rate of range dynamics on the other, may occur at roughly the same time scale, raising questions about the role of spatial dynamics in evolutionary radiations, and indeed questions about interactions between speciation–extinction dynamics and the dynamics of geographic ranges. This is an example of eco-evolutionary dynamics during a very long period of time.

Phylogenies are routinely used to gain insight into the processes that have shaped extent diversity. Intuitively, a decrease in species accumulation rate through time would signal adaptive radiation. To formally test such idea, one approach is to analyse the internode distance between speciation events (Pybus & Harvey 2000). The approach relies on comparing the  $\gamma$ -statistic from the empirical phylogeny, which summarizes the distribution of internode distance, with the distribution of  $\gamma$ -statistics obtained from a null model of constant speciation and extinction rates. A significant statistical difference between the empirical and the null distribution is then used as an evidence of adaptive radiation (Pybus & Harvey 2000). Another approach assesses which model fits the data better: a model with constant or with negative diversity-dependent diversification rate (Etienne & Haegeman 2012, Etienne *et al.* 2012). A better fit of the latter (e.g., with lower AIC) serves as an evidence of adaptive radiation. The advantage of the second approach is that it also estimates speciation and extinction rates, and the carrying capacity. The major weakness is that both approaches ignore

species' traits (e.g., geographical range) which might also affect diversification rate.

To allow extinction and speciation events to depend on species' traits, Maddison *et al.* (2007) developed the Binary-State Speciation-Extinction (BiSSE) model. In practice, the BiSSE model is of limited use while considering geographic range as species' traits because it constrains each lineage to possess only one character. To tackle that limitation, Goldberg *et al.* (2011) developed the Geographic-State Speciation-Extinction (GeoSSE) model which assumes the geographic range as a species trait and incorporates explicit spatial dynamics coupled with the diversification process. The GeoSSE model captures the facts that species' ranges may influence speciation and extinction rates, and *vice versa*, via between-region speciation and regional extinction of widespread lineages. To date, the GeoSSE is perhaps the most flexible model of macroevolution that incorporates range dynamics, and it has been used for a better understanding of patterns of diversification and species' geographic distributions (Buerki *et al.* 2012, 2013, Bloom *et al.* 2013, Jansson *et al.* 2013, Rolland *et al.* 2014).

One major limitation of the GeoSSE however is that it only permits comparative estimates between two regions. Here, we extend the GeoSSE model to allow for multiple regions and hence a more elaborate description of range dynamics. In addition, we also implement diversity (species number) dependence in the rate of speciation. This last assumption is new since previous models assume either constant rates of speciation, extinction, and dispersal (Stadler 2013), or diversity-dependent rates but without range dynamics (Rabosky & Lovette 2008, Etienne & Haegeman 2012). Instead of using the likelihood-based approach, we opt for Approximate Bayesian Computation (ABC), which allows inference of more flexible models and avoids the computationally intensive method of solving the likelihood function (Toni *et al.* 2009, Beaumont 2010, Sunnaker *et al.* 2013). We aim to estimate the parameters of two modes of speciation i.e., between-region (allopatric) and within-region speciation, as well as extinction, regional colonization, and the initial range of the last common ancestor (thereafter initial range).

We apply the model to an evolutionary radiation of dung beetles in Madagascar. Madagascar has a unique biota (Myers *et al.* 2000) with an exceptionally high level of endemism: 100% in Amphibians and terrestrial mammals, 92% in reptiles, 90% in plants, 44% in birds (Goodman & Benstead 2003), and 96% in dung beetles (Miraldo *et al.* 2011). The obvious cause of high endemism even at the family level (Vences *et al.* 2009) is the ancient isolation of Madagascar, which became separated from the African continent around 135 million years ago and from the Indian plate around 90 million years ago (de Wit 2003). In the case of dung beetles,  $\beta$ -diversity (spatial turnover in the species composition) is significantly higher in Madagascar than in other tropical regions (Viljanen *et al.* 2010), indicating high frequency of species with small geographical ranges. The mechanisms that have been suggested to be responsible of microendemism include ecogeographic constraints, isolation of taxa in rainforest refugia during climatically dry periods, riverine barriers to dispersal, and montane refugia (Vences *et al.* 2009). Undoubtedly, the large size of Madagascar (587 000 km<sup>2</sup>), the extensive mountain range from north to south, and the diversity of dissimilar biomes have all played a role in the evolution of the Malagasy biota (Wilme *et al.* 2006, Yoder & Nowak 2006, Vences *et al.* 2009). There are around 300 extant species of dung beetles, which appear to originate from eight independent colonizations (Wirta *et al.* 2010, Miraldo *et al.* 2011). In this paper, we focus on the most recent and most successful (in terms of net speciation rate) radiation, consisting of the genera *Nanos* and *Apotolamprus* with 74 known species (Miraldo *et al.* 2011, Miraldo & Hanski 2014).

We start by describing the mathematical model and the lineage of Malagasy dung beetles to which we apply it. Next, we describe the statistical model with details of the ABC implementation. We describe how the model is simulated and validated using data with comparable structure to the empirical data. We could not reliably estimate the full model with region-specific parameters of both speciation and extinction most likely due to parameter correlations. Assuming equal extinction rate in each region yielded informative posterior distributions for

the region-specific speciation and colonization parameters. In the empirical data, the estimated between-region colonization and within-region extinction rates are of similar magnitude. We conclude that the most significant difference between the regions is exceptionally high diversification rate in northern Madagascar, where the current species richness is highest.

## Material and methods

### Model

The model, which we call the GR (Geographic-Radiation) model for short, is an extension of the GeoSSE model (Goldberg *et al.* 2011) to multiple regions. The GR model is a continuous-time discrete-space Markov process that models the occupancy state of each region with no consideration for population sizes in the regions. While modelling the dynamics of a lineage, three events may occur: colonization, extinction, and speciation. Colonization can only happen between adjacent regions, at a per-lineage rate  $c$ , and it turns an unoccupied region into an occupied region. An occupied region may become unoccupied due to an extinction event, which happens in each region at a per-lineage rate  $e$ . An extinction event leads to either range contraction, if the lineage occurred in two or more regions, or to global lineage extinction, if the lineage was endemic to the region. We assume throughout this paper that  $c$  is the same for all regions, while  $e$  is either the same for all regions or has region-specific values.

Speciation may happen in two different ways, either within a region or between regions. Within-region speciation ( $s$ ) refers to a speciation event within a single region. Within-region speciation event gives rise to two sister species, the first one having the same distribution as the parent species and the second one being endemic to the region in which speciation occurred. We always assume that  $s$  has region-specific values. In between-region (allopatric) speciation, the ancestral species occupies at least two regions, and the ranges of the sister species are obtained by splitting the ancestral species' range into two disjoint sub-regions. If spatial configura-

tion is one-dimensional, as it is here (below), and if the range of the ancestral species consists of  $m$  regions, there are  $m - 1$  possible bipartitions, each of which leads to between-region speciation at a constant rate  $a$ . Hence, the wider the distribution of a lineage, the greater the probability of between-region speciation. Finally, to initiate the diversification process, we need to specify the initial range  $i_R$  which gives the range (a single region) of the last common ancestor.

The number of extant species in a region can affect the dynamics and thereby introduce diversity (lineage number) dependence into the model. For simplicity, we assume the same functional form for each region and that only within-region speciation rate is diversity-dependent. We model this by multiplying the per-lineage within-region speciation rate by a decreasing function  $f(n) = 0$  if  $n = 0$  and  $10/n$  otherwise, where  $n$  is the number of lineages in the region. The inverse function is chosen to balance two mechanisms: large number of species lowers the per lineage speciation rate (e.g., because of competition), but also increases the pool of genetic diversity. Hence, the total within-region speciation rate remains constant in each region. This functional form has been previously used for time-dependent speciation rate (Nee *et al.* 1994a). Diversity-dependence can take other functional forms and may include an additional free parameter such as the carrying capacity or the rate of decline (e.g., exponential) for each region (Rabosky & Lovette 2008, Etienne & Haegeman 2012). However, we prefer not to parametrize the functional form in our model to facilitate the inference, because it already has a large number of region-specific parameters (*see below*).

### Application to Malagasy dung beetles

We apply the model to the radiation consisting of the genera *Nanos* and *Apotolamprus*, which is known to be monophyletic and colonized Madagascar 24 to 13 million years ago (Miraldo & Hanski 2014). There are altogether 74 known species (Montreuil *et al.* 2014), of which 50 species are included in the molecular phylogeny of Miraldo and Hanski (2014). Miraldo and Hanski (2014) further showed a decline in the diversifi-

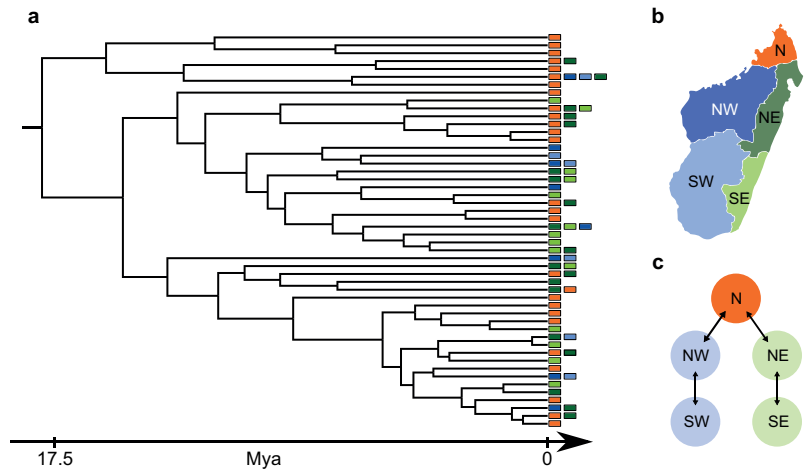
cation rate toward the present, which makes the data suitable for modelling purposes. Madagascar is divided into five geographical regions: North (N), Northeast (NE), Northwest (NW), Southeast (SE), and Southwest (SW) (Fig. 1b and c). The regions are not physically isolated but they have different climatic and environmental conditions and correspond to a high-level classification of recognized Malagasy biogeographic regions (Wilme *et al.* 2006). There is an especially strong contrast between the humid eastern regions (with rain forest) and dry western regions (various dry forest types and open habitats). Considering this contrast, and the fact that the high-altitude plateau in central Madagascar has very few dung beetles, colonization between East (NE and SE) and West (NW and SW) is very unlikely and we set the respective colonization rates to 0. The spatial configuration of the five regions is thereby reduced to a one-dimensional space (Fig. 1c). Below, when we allow for variation in speciation and extinction rates among the regions, we assume that the two eastern and the two western regions share the same values, thus there may be three region-specific speciation ( $s_W, s_N, s_E$ ) and extinction ( $e_W, e_N, e_E$ ) rates.

### Statistical model

Our goal is to infer the distribution of the model parameters that would produce the empirically observed phylogeny and the pattern of geographical ranges of the species (Fig. 1a). We consider a full and a simplified scenario to explore the power of the model in inferring the parameters. In the full scenario, we attempt to estimate both region-specific speciation and extinction rates, and thus the vector of parameter values is given by  $\phi_f = (s_W, s_N, s_E, e_W, e_N, e_E, c, a, i_R)$ . In the simplified scenario, we assume that all extinction rate parameters have the same value ( $e = 1$ ), and we thus estimated  $\phi_s = (s_W, s_N, s_E, c, a, i_R)$ .

The closely-related GeoSSE model (Goldberg *et al.* 2011) assumes two regions, for which the likelihood function is constructed and solved. This leads to a set of three differential equations along each branch of the tree. In this approach, the number of differential equations increases exponentially with the number of regions, and

**Fig. 1.** (a) Phylogeny and distribution of the *Nanos–Apotolamprus* clade (adapted from Miraldo & Hanski 2014). Each color corresponds to a different region in Madagascar (see map in b). (c) Geographical setting of the five regions with their respective connectivity represented by arrows.



becomes intractable if the model incorporates diversity-dependent parameters. Therefore, instead of constructing and solving the likelihood function required by traditional Bayesian computational methods (Robert & Casella 2004), we employ Approximate Bayesian Computation (ABC) to estimate the posterior probability density of the parameters. In the application to Malagasy dung beetles below, we assume uninformative uniform prior distributions for  $\varphi$ , with ranges [0, 10] for the extinction and within-region speciation rate parameters, [0, 5] for the between-region speciation rate parameter, and [0, 2] for the colonization rate parameter.

ABC is a flexible family of methods based on simulating an appropriate model. In the last decade, ABC has proliferated due to the increasing complexity of data and models that researchers study (Toni *et al.* 2009, Beaumont 2010, Sunnaker *et al.* 2013). ABC is a helpful method when the likelihood function is too costly to compute or is intractable (which is the case here). There are several versions of ABC, but the main idea is to (1) generate a vector  $\varphi$  of parameter values from a distribution, usually the prior distribution, (2) simulate the mathematical model using  $\varphi$  to generate a simulated data set, and (3) compare the simulated and empirical data sets. If the distance between the two data sets is smaller than a fixed tolerance  $\varepsilon$ , the proposed set of parameters is accepted. If these steps are repeated many times and the tolerance  $\varepsilon$  is small enough, meaning that we only accept

parameters that produce a data set that is very similar to the empirical one, the set of accepted parameter values provides a good approximation of the true unknown posterior distribution of the model parameter. In the rest of this paper, we use interchangeably posterior distribution and approximate posterior distribution. ‘Distance’ is interpreted liberally, and if the data cannot be compared directly, a set of summary statistics is used to capture the essential features of the data. The summary statistics project the data into a low-dimensional space and introduce another level of approximation to the method. Nonetheless, with sufficient summary statistics, the distribution targeted by the ABC algorithm matches the posterior distribution, but such choices are generally unavailable outside simple models. In general, the summary statistics have to be constructed by balancing the amount of information they carry with low dimensionality to keep the computational complexity of the algorithm feasible. The choice of summary statistics is therefore a crucial component of ABC, and we next describe the summary statistics used here, followed by a detailed description of the implementation of the ABC.

## Summary statistics

As explained above, summary statistics are the key components of ABC. We tested several statistics such as tree imbalance described by the



Colless index (Colless 1982), the distribution of relatedness (i.e., how many internal nodes need to be crossed to move from one tip to another, with sister and cousins species having relatedness one and three, respectively), and the Lineage-Through-Time (LTT) plot but none of these were informative. At the end, we retained four summaries: the number of species per region, the distribution of species' range sizes, the distribution of range similarity between pairs of sister species, and the relative phylogenetic diversity. We did not include tree size as summary statistics but instead use it as an initial criterion to accept computing the actual statistics (steps 3 and 10 in the algorithm below). The tolerance of the criterion is done by setting an upper limit on the size of the simulated tree ( $n_{\max}$  below). Moreover, it is essential to compare trees of the same size (the same number of tips). Since the number of species obtained at the end of each simulation will most likely be different from the number of species in the empirical tree, denoted by  $n_e$  which itself is smaller than the true unknown total number of species, we randomly sample  $n_e$  species from the simulated data and prune the resulting tree accordingly. The simulation parameters are described below. The distance between the simulated and empirical trees, denoted by  $A$  and  $\hat{A}$  respectively, is computed using the distances based on the four summaries.

In the case of the number of species per region, we calculate the Euclidian distance between the two trees as

$$d^1 = \sqrt{\sum_{i \in \{N, NE, NW, SW, SE\}} [s(A, i) - s(\hat{A}, i)]^2}, \quad (1)$$

where  $s(\cdot, i)$  counts the number of species in region  $i$ . The second distance  $d^2$  compares the distributions of range sizes, and is obtained by computing the difference between the number of species occupying 1, 2, ..., 5 regions,  $r(\cdot, i)$ , as

$$d^2 = \sqrt{\sum_{i \in [1, 5]} [r(A, i) - r(\hat{A}, i)]^2}. \quad (2)$$

The third distance measures the similarity in the distributions of sister species. We consider three categories of distributions. If the sister species have exactly the same distributions their pairwise distance is 0. If the distribution of one species is nested within the distribution of the

other one the distance is one, and otherwise the distance is two. Our third summary distance is then given by

$$d^3 = \sqrt{\sum_{k=0}^2 [u(A, i) - u(\hat{A}, i)]^2}. \quad (3)$$

Here  $u(\cdot, k)$  computes the proportion of sister species pairs that have the same distributions ( $k = 0$ ) have nested distributions ( $k = 1$ ), and other types of distributions ( $k = 2$ ). Finally, phylogenetic distance was calculated by summing up the total length of the branches (starting from the crown age) and divided by the crown age. This function, denoted by  $v(\cdot)$ , allows one to characterize the relative phylogenetic diversity of any tree, regardless of the length of the simulation. The fourth distance is given by

$$d^4 = |v(A) - v(\hat{A})|. \quad (4)$$

## Implementation of ABC

We use a sequential Monte Carlo ABC (SMC-ABC) combined with adaptive tuning of tolerance and adjustment of the perturbation kernel which is comparable to the framework used by Numminen *et al.* (2013). In sequential ABC, instead of using one value for tolerance and one round of rejecting samples, a sequence of intermediate distributions with decreasing tolerances are utilized. Each generation is terminated when a required number of parameter values  $N$  has been accepted. The set of accepted parameters  $\{\varphi_i\}_{i=1, \dots, N}$  and their associated weights  $\{w_i\}_{i=1, \dots, N}$  represent a sample from the intermediate distribution, which is used as a proposal distribution for sampling in the next generation. To ensure that sufficient variation remains among the parameters in successive generations, the parameters are perturbed according to a perturbation kernel. Because we have both continuous and categorical variables to estimate, the perturbation kernel  $\mathbf{K}$  is a Cartesian product of a multivariate normal distribution  $\mathbf{K}_1$  and a circular kernel  $\mathbf{K}_2$ . The variance-covariance matrix  $\Sigma$  for kernel  $\mathbf{K}_1$  is computed based on the sample from the previous generation

$$\sum_{i,j} = \sum_{k=1}^N \sum_{l=1}^N w_k w_l (\varphi_l^i - \varphi_k^i) \cdot (\varphi_l^j - \varphi_k^j). \quad (5)$$

Kernel  $\mathbf{K}_2$  is used for the initial range and is assumed to be circular on the set  $\{\text{SW, NW, N, NE, SE}\}$  with a probability  $p/2$  to move to the right,  $p/2$  to move to the left, and  $1-p$  to remain in the same region.

The vector of tolerance values  $\boldsymbol{\varepsilon} = (\varepsilon^1, \varepsilon^2, \varepsilon^3, \varepsilon^4)$  is updated in the following manner. If  $\{(d_i^1, d_i^2, d_i^3, d_i^4)\}_{i=1, \dots, N}$  is the distance between the simulated and empirical data for the accepted parameter vector  $\{\varphi_i\}_{i=1, \dots, N}$  then  $\varepsilon^j$  for the next generation is chosen as the quantile of  $d^j$  such that a fixed percentage of  $d^j$  is accepted. This provides a robust way of combining information from the different statistics and accounts for possible correlations between the statistics.

The exact algorithm is as follows:

#### A. First generation:

1. sample a proposed parameter vector  $\boldsymbol{\varphi}$  from the prior distribution  $\boldsymbol{\pi}$ ,
2. simulate a tree using parameter values  $\boldsymbol{\varphi}$  and the GR model (for the length of the simulation *see* below),
3. if the number of species is smaller than  $n_{\min}$  or greater than  $n_{\max}$  go to step 1, otherwise accept  $\boldsymbol{\varphi}$ ,
4. sample  $n_c$  species from the simulated tree and compute distance  $\mathbf{d} = (d^1, d^2, d^3, d^4)$  between simulated and empirical trees,
5. if the number of accepted parameter values (= size of the sample) is strictly smaller than  $N$ , go to step 1, otherwise set  $w_i = 1/N, i = 1, \dots, N$  and go to B.

#### B. Subsequent generations:

At this stage, we have a set  $\{\varphi_i, w_i, d_i\}_{i=1, \dots, N}$  representing the set of accepted parameters, their associated weights, and the distance between the simulated and empirical data. The next steps are:

6. compute the variance-covariance matrix  $\boldsymbol{\Sigma}$ ,
7. initialize the tolerance  $\boldsymbol{\varepsilon} = (\varepsilon^1, \varepsilon^2, \varepsilon^3, \varepsilon^4)$ ,
8. sample  $\boldsymbol{\varphi}^*$  from  $\{\varphi_i\}_{i=1, \dots, N}$  with respect to  $\{w_i\}_{i=1, \dots, N}$  and perturb such that  $\boldsymbol{\varphi} \sim \mathbf{K}(\boldsymbol{\varphi}^*, \boldsymbol{\Sigma}, p)$ ,
9. simulate tree using the perturbed values  $\boldsymbol{\varphi}$  and the GR model,

10. if the number of species is smaller than  $n_{\min}$  or greater than  $n_{\max}$  go to step 8,
11. sample  $n_c$  species from the simulated tree and compute distance  $\mathbf{d}$  between simulated and empirical trees,
12. if the distance  $d^i \leq \varepsilon^i$  for each  $i = 1, \dots, 4$  accept  $\boldsymbol{\varphi}$ ,
13. if the number of accepted parameter values is smaller than  $N$ , go to step 8, otherwise go to C.

#### C. For each $k = 1, \dots, N$ , compute

$$w_k \propto \frac{\pi(\varphi_k)}{\sum_{n=1}^N \widetilde{w}_n K(\widetilde{\varphi}_n, \varphi_k, \boldsymbol{\Sigma}, p)},$$

and normalize such that  $\sum_{k=1}^N w_k = 1$ . Here  $\{\widetilde{\varphi}, \widetilde{w}\}$  denotes the vector of parameters and weights from the previous generation (i.e., the proposal distribution), whereas  $\varphi_k$  is the newly accepted parameter vector. Return to B if the resulting distribution has not converged. For convergence, we rely on visual inspection of the marginal posterior distributions.

### Model simulations

The simulation of the GR model is based on the Gillespie algorithm; the simulation and the ABC algorithm are implemented in Mathematica ver. 9.0 (Wolfram Research Inc.). We initialize the simulation with one species occurring in a single geographic region, and the simulation is run until time  $T$ , and the parameter vector was rejected if the number of species was not within the required limits (*see* below). Throughout the manuscript, we set  $T = 16$  but we report additional results for  $T = 8$  in the supplements. The values represent either a late phase, when the number of species has already fluctuated around the expected value for a long time, or an early phase of the radiation, when the quasi-stationary equilibrium has been recently reached. In the GR model, the absolute value of  $T$  is arbitrary as the other parameters can always be scaled (e.g., by multiplying  $f$  by a constant or by assigning a different fixed extinction rate in the simplified model). The main goal is to compare multiple diversification rates for multiple regions.



As explained above, there are 74 described species in the *Nanos–Apotolamprus* radiation, but only 50 species were available for the molecular phylogeny (Miraldo & Hanski 2014). Additionally, we excluded one of the 74 species because it is known from a single specimen from a single locality. We require that the number of species at the end of the simulation is greater than  $n_e = 73$  and smaller than  $n_{\max}$ . If this condition is met, we obtain two samples from the simulation. First, we sample 73 species from the simulation and compute  $d^1$  and  $d^2$ , after which we sample 50 species from the sample of 73 species and compute  $d^3$  and  $d^4$ . For the first generation, we accept the parameter vector as long as the number of species is between  $n_e = 73$  and  $n_{\max} = 150$ .

## Inference validation

We validated our approach to parameter estimation by applying the statistical model to simulated data. For the full and the simplified scenarios, we generated respectively 10 and 20 random vectors of parameter values  $\phi$  based on the priors used in the empirical case. For practicality, the simulated data needs to abide by three criteria. First, the parameters cannot be too close to the upper boundary of the prior distribution, as the posterior distribution with the data sets generated using extreme values would be limited by the prior support. In practice, we generated parameters within the lower three fourths of the range of the prior. Second, we selected parameters that resulted in at least 50 species and at most 150 species. Third, each parameter should consistently produce trees of similar size. In practice, we only accepted parameter values when the difference between the first and two out of four other replicates was less than 20. We then estimated each set of parameters using the ABC framework described above, requiring that  $n_{\min} = n_s$  and  $n_{\max} = n_s + 50$  where  $n_s$  is the total number of species in the first generated data. We estimated each parameter using the same simulation time  $T$  that was used to generate the data.

We evaluated the performance of the ABC estimation procedure by calculating the uncertainty, error and bias of the parameter estimates.

The uncertainty was assessed by calculating the mean and the 95% posterior interval of the marginal approximate posterior distribution. To measure the error and bias of the parameter estimates, we first calculated the mean squared error for parameter  $\phi^i$  as

$$\text{MSE}(\phi^i) = \sum_{k=1}^N w_k (\phi_k^i - \phi_T^i)^2,$$

where  $w_k$  is the weight associated with  $\phi_k$  and  $\phi_T^i$  is the true parameter value used for simulating the data set. The error associated with each parameter estimate was now measured with the root mean squared error

$$\text{RMSE}(\phi^i) = \sqrt{\text{MSE}(\phi^i)}.$$

The bias associated with each parameter was obtained by using the property  $\text{MSE}(\phi^i) = \text{var}(\phi^i) + [E(\phi^i) - \phi_T^i]^2$ , where  $E(\phi^i)$  and  $\text{var}(\phi^i)$  are the posterior mean and variance of  $\phi^i$  respectively. We obtain a measure of bias as

$$\text{bias}(\phi^i) = \sqrt{\text{MSE}(\phi^i) - \text{var}(\phi^i)}.$$

For the discrete variable of initial range, we computed the probability of having the correct value.

Finally, to explore the effect of tree size on the performance of the model in the simplified scenario, we generated 10 trees of different sizes while keeping the ratio of within-region speciation rates constant. We chose four cases where the ratio  $s_w:s_E:s_N$  are 1:1:1, 1:1:4, 1:4:8, and 4:4:1. We then calculated the scaled root mean squared error which is RMSE divided by the true value.

## Results

### Inference validation

In the full scenario, we attempted to estimate a model in which there are three region-specific extinction rate parameters,  $e_w$ ,  $e_N$ , and  $e_E$ , and three region-specific within-region speciation rate parameters,  $s_w$ ,  $s_N$ , and  $s_E$ , one colonization rate  $c$ , one between-region speciation rate  $a$ , and the initial range  $i_R$ . The values of the model parameters were poorly estimated. There was

high uncertainty in parameter estimates, and small values were systematically overestimated (Fig. 2). A likely cause of the relatively poor inference was due to strong correlations between extinction and speciation rate parameters within each region (Table 1). The next strongest correlations were between colonization and extinction rates (All pairwise correlations are shown in Table A1\_1).

In the simplified scenario, we fixed all extinction rates to 1 (the results for  $T = 16$  are shown in Fig. 3). The within-region speciation rates were relatively well estimated and the bias associated with the estimates was low (Fig. 3). The error of the estimates, measured by the RMSE, was mostly low for  $s_w$  and  $s_E$ , but higher for  $s_N$ . The latter result is expected as the information comes from a single region and is hampered by the stronger effect of colonization (Fig. 1c). In contrast with the full scenario, the error and bias in the simplified scenario are small and increase slightly with decreasing value of the true parameter (Fig. 2: gray vs. black lines in the bottom row). The colonization parameter  $c$  was estimated similarly well as  $s_w$  and  $s_E$ , although the error was more due to bias (underestimation) than posterior uncertainty (2nd row in Fig. 3). Overestimation was most apparent for extremely small values of the parameters (two orders of magnitude smaller) for the within-region speciation and colonization rates which led to large RMSE and consequently the bias. The overestimation could also be related to the effect of the prior distribution in shifting the estimate towards the prior expectation.

The between-region speciation rate ( $a$ ) was not identifiable with the method, and the estimate showed both high uncertainty and bias (4th column in Fig. 3). The results were qualitatively similar for a different value of  $T$  (Fig. A1\_1) but the inference worked best when  $T$  is large probably due to reduced transient in the simulated trees. Inferring the region of the initial range turned out to be difficult. The probability of correctly estimating the initial region was around 0.2 (Fig. A1\_2), which is the probability of randomly choosing a region out of five. However, for some parameter values the probability of correctly estimating the initial region was as high as 0.4 (Fig. A1\_2). This happened when the

initial region was either SW or SE, the two most peripheral regions (Fig. 1c).

In the simplified scenario, we performed an additional test to assess the power of the estimation with respect to the size of the tree. The general conclusion is that the scaled root mean squared error decreases with increasing number of species (Fig. A1\_3). Thus, the parameters were better estimated for larger trees.

Application to empirical data

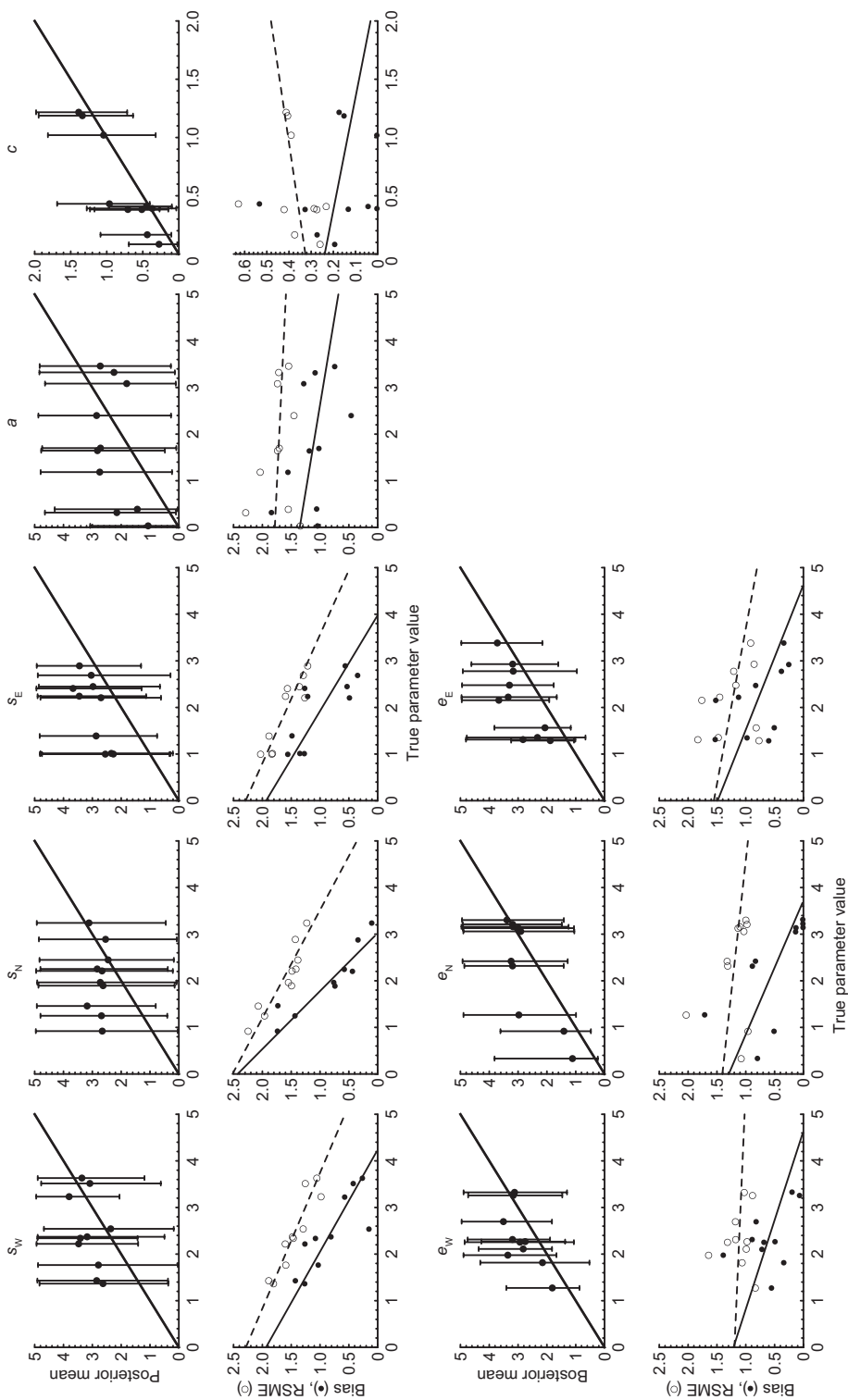
Since inference validation works best in the simplified scenario, we used that scenario for comparing within-region speciation rates while assuming extinction rates are equal to one. The rate of between-region speciation had a mode close to zero while the bulk of the distribution lay between 0 and 2 (Fig. 4). This parameter could not be estimated in the simulated data (Fig. 3). Due to the diversity-dependence, comparing estimated within-region speciation, with other process is not straightforward. However, considering that we fixed extinction rates to 1, the results indicate that colonization happened at a lower rate (mean around 0.4; see Fig. 4). Although the model tended to underestimate colonization rate in the simulated data, even the upper limit of the distribution of the colonization rate is smaller than 1. More importantly, there is a clear difference among the regions: within-region speciation rate was clearly higher in North than in East and West (Fig. 4).

Estimating the initial region  $i_r$  was clearly difficult as the posterior distribution resembles a uniform distribution irrespective of the scenario

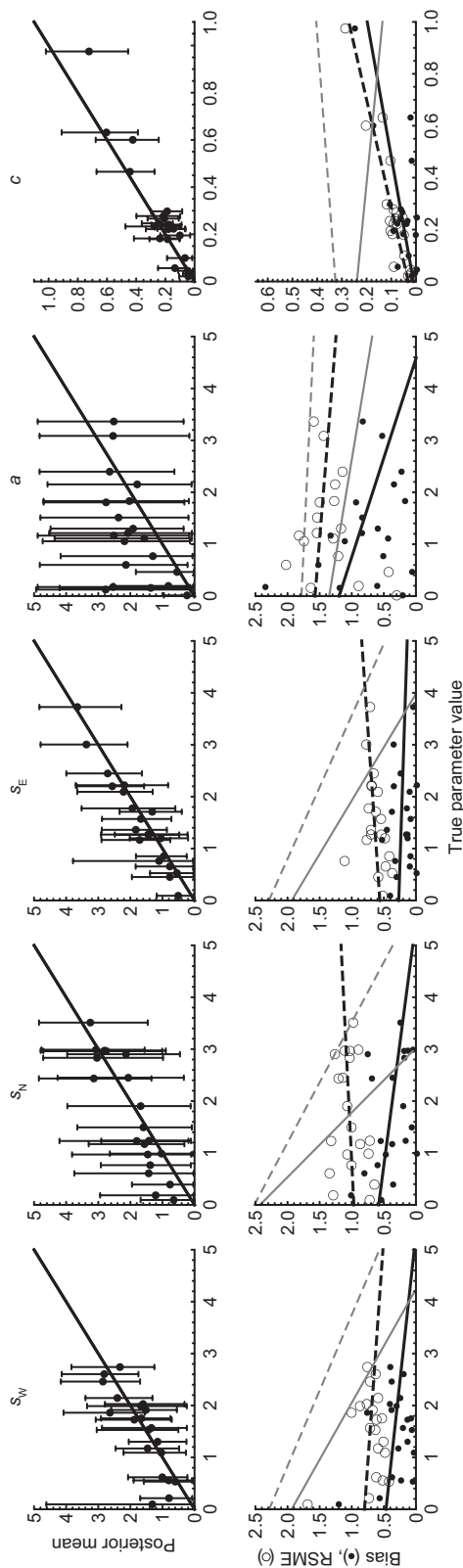
Table 1. Highest correlations between parameters in the full scenario.

Parameter 1	Parameter 2	Correlation*
$s_w$	$e_w$	0.458
$s_E$	$e_E$	0.456
$s_N$	$e_N$	0.284
$c$	$e_E$	0.262
$c$	$e_w$	0.259
$c$	$e_N$	0.221

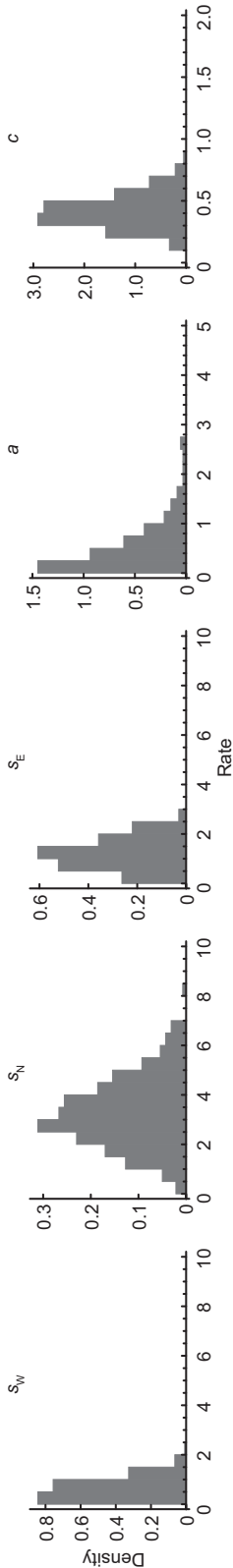
\* Average Pearson's correlation of 10 simulated data sets.



**Fig. 2.** Inference validation in the full scenario for 10 simulated data. For the first pair of rows, columns from left to right denote respectively within-region speciation rate in the West ( $s_w$ ), the North ( $s_N$ ), the East ( $s_E$ ), between-region speciation rate ( $a$ ), and colonization rate ( $c$ ). For the second pair of rows, columns from left to right denote respectively extinction rate in the West ( $e_w$ ), the North ( $e_N$ ), and the East ( $e_E$ ). The 1st and 3rd rows show the true values (x-axis) against the means of the marginal approximate posterior distribution (dots) and their 95% posterior intervals (error bars). The 2nd and 4th rows depict the true parameter values (x-axis) against the root mean squared errors (RMSE, y-axis) and the roots of the difference between the mean squared errors and the posterior variances (bias, y-axis). Dashed and solid lines are fitted linear regressions of the RMSE and bias, respectively. Simulation parameters:  $T = 16$ ,  $N = 400$  and  $q = 50\%$ .



**Fig. 3.** Inference validation in the simplified scenario for 20 simulated data, in which extinction rate is fixed to 1 for all regions. Columns from left to right represent respectively: within-region speciation rate in the West ( $s_w$ ), the North ( $s_n$ ), and the East ( $s_e$ ), between-region speciation rate ( $c$ ). The upper row shows the true values (x-axis) against the means of the marginal approximate posterior distributions (dot) and their 95% posterior intervals (error bars) (y-axis). The lines represents the identity line. The bottom row depicts the true parameter values (x-axis) against the root mean squared errors (RMSE, circle, y-axis) and the roots of the difference between the mean squared errors and the posterior variances (bias, dot, y-axis). Black dashed and solid lines are fitted linear regressions of RMSE and bias, respectively. For comparison, fitted regressions from Fig. 2 are shown as gray dashed (RMSE) and solid (bias) lines. Simulation parameters:  $T = 16$ ,  $N = 400$  and  $q = 50\%$ .



**Fig. 4.** Approximate posterior distributions for the empirical data in the simplified scenario. Columns from left to right represent respectively: within-region speciation rate in West ( $s_w$ ), North ( $s_n$ ), and East ( $s_e$ ), allopatric speciation rate ( $a$ ), and colonization rate ( $c$ ). The distributions are from the 7th generation. Parameters:  $q = 50\%$  and  $N = 1000$ .

and the length of the simulation (for complete results *see* Figs. A2\_1–2). The empirical data suggests rather strongly that the initial colonization occurred in North (Miraldo & Hanski 2014). We hence repeated the estimation of the other parameter values after fixing the initial range to North and we found no substantial differences between the estimates for the other parameters (Fig. A2\_3).

## Discussion

In this work, we have developed a model to infer and compare diversity-dependent speciation and extinction, and colonization rates for multiple regions. Our geographical radiation model (GR) is an extension of the GeoSSE which is limited to two regions and does not allow diversity-dependent parameters (Goldberg *et al.* 2011). Models used for phylogenetic analyses are often limited by the ability of researchers to solve them, which favours relatively simple models. Although general understanding of basic processes may be obtained with simple models [e.g., the Yule model (Yule 1924)], biologists remain keen to consider models with more complex assumptions. The constraint of using simple model and thus the ability to write down the likelihood function can be alleviated by using ABC. Here, ABC can be helpful in striking a balance between model complexity and analytical or computational tractability. Instead of solving the likelihood function as in the SSE models (Maddison *et al.* 2007, Goldberg *et al.* 2011), the ABC approach allows us to construct models based on biologically well-justified assumptions.

The flexibility of ABC is obtained by targeting an approximate version of the posterior distribution, instead of the exact one. A crucial component in controlling the degree of approximation in ABC algorithms is the choice of the summary statistics that project the data to a lower dimensional space without losing too much information about the parameters of interest. While there exists a range of methods for reducing the dimensionality of candidate statistics (Blum *et al.* 2013), the performance of these statistics may still be limited, because the methods seek for simple global relationships between the summary

statistics and parameters. Therefore, choosing the summary statistics based on careful inspection of the model structure and observed data, as we have done here, is important.

Adaptive ABC algorithms such as SMC-ABC that we used allow efficient exploration of the parameter space. These algorithms require little prior information on plausible parameter values, as they automatically detect regions of high posterior density. This makes them suitable for models with a large number of parameters, where rejection sampling-based techniques suffer from the curse of dimensionality. However, because they target posterior distributions associated with one particular data set, validating the inference procedure poses a considerable challenge and requires substantial computational effort. This is in contrast with the rejection sampling-based algorithms, where the same simulated parameters and data sets can be used to analyse multiple test data sets.

We expected that we would not be very successful in inferring the parameters of region-specific speciation and extinction rates simultaneously (full scenario) because of parameter correlations, which turned out to be the case. Goldberg *et al.* (2011) found the highest correlation between extinction and colonization rate parameters, and the next highest between speciation and colonization rates. In contrast, here the highest correlations were between region-specific within-region speciation and extinction rates (Tables 1 and A1\_1). Such correlation makes sense because GR assigns different speciation and extinction rates for each region in contrast with a single parameter for colonization rate for all regions. The BiSSE model, which is closely related to our model and the GeoSSE, reported confounding factors that limit the power of the model (Davis *et al.* 2013), especially when the ratio between the observed characters is highly skewed. Likewise, we found that poorly estimated parameters are associated with a low parameter value in colonization, extinction or within-region speciation rates in particular regions, which leads to skewed distribution of the number of species in each region.

A central question is the cost of using ABC instead of solving exact likelihood function for the statistical inference. We may not have maxi-

mally used information about the timing of the tree as the internode distance is essential for estimating both speciation and extinction (Nee *et al.* 1994b). In general, the failure to infer all parameters might be caused by non-sufficient statistics. Our approach cannot infer the absolute tempo of diversification but is best for comparing relative diversification rate among regions which is often the main interest. The amount of information lost can be assessed by comparing the estimates from a simplified version of GR (with 2 regions and without diversity-dependence) and GeoSSE. We also suspect that even if the statistics are sufficient — and thus the inference will be equivalent to solving the likelihood — the amount of information in a phylogeny can be insufficient to answer the more detailed questions and to infer large number of parameters. A further set of hierarchical models to differentiate the effect of adding the number of regions and the role of diversity dependence would also reveal the strengths and the weaknesses of the approach but we leave that for future studies.

A more serious issue that concerns all comparative methods for discrete character states was recently reported. The problem is that phylogenetic comparative method for discrete character states returns high rate of false positive (Madison & FitzJohn 2015). Rabosky and Goldberg (2015) showed that even if trait change is neutral, and thus randomly distributed on a tree, BiSSE still associates a high speciation rate to the most frequent character state. The origin of the problem is still unknown but a potential fix is to test for model adequacy which is available for continuous but not for discrete character states (Pennell *et al.* 2015). In this paper, we underscore some summary statistics that might be useful for testing adequacy in the discrete case.

The approach we use here provides several practical advantages compared to the other models. First, hardly all taxa are included in the phylogeny. Some methods circumvent the incomplete taxa sampling by deriving an appropriate likelihood function (FitzJohn *et al.* 2009). GR offers an alternative solution, as our forward simulation mimics the actual radiation, we can simply prune the tree at the end of the simulation before computing the summary statistics. We used random sampling but it is trivial to choose



other sampling methods — say the pruning depends on range size. Second, unlike GeoSSE and most State-Speciation-Extinction models which require a dated phylogeny, the GR model can be applied to non-ultrametric trees. A final important advantage of the GR model is that it uses information (e.g., on geographical ranges) on species that are not included in the phylogeny. In the present case, though only 50 species out of the 73 species are included in the phylogeny, we have used data on geographical ranges for all the 73 species in the analysis.

Returning to our empirical results, between-region speciation rate and the initial range, where the colonization of Madagascar originally took place, could not be estimated in the simulated data and the estimate in empirical data is therefore not reliable. Goldberg *et al.* (2011) similarly found that the between-region speciation rate had the broadest posterior distribution. When they fixed the within-region speciation rate to a constant, the estimation of the between-region speciation rate was more successful. Others have concluded that inferring geographical mode of speciation is difficult because fast range dynamics may dilute the relevant signal (Losos & Glor 2003). It is also possible that between-region speciation rate was difficult to estimate here because of the paucity of information in the data, given the five regions and the many ways a large geographical range can be split into two. With simulated data, the initial range was best estimated when simulation time was short and the initial range was SW or SE, where the current number of species is lowest. In other cases, we conjecture that high rate of within-region speciation may produce convergent patterns and reduce the influence of where the process initially started.

We found that the colonization rate from one region to another was intermediate in magnitude to extinction rate in the different regions. This result clearly highlights the importance of integrating the study of range dynamics with the study of speciation–extinction dynamics in evolutionary radiations across large area. This result is also consistent with the current geographical distributions of the species. For instance, in a clade of 24 large-bodied *Nanos* species out of the 74 species studied here, Miraldo and Hanski (2014) found that the species have

largely allopatric distributions, with different species groups occurring in different parts of Madagascar. Such a pattern can arise only if the species ranges are very conservative, that is, if the region-to-region colonization and diversification processes occur at the same rate. This is an example of eco-evolutionary dynamics occurring at evolutionary time scale.

The simplified scenario studied here produced a clear relationship among the modes of the distribution of region-specific parameter. Speciation rate is clearly highest in North than in West or East where the current species number is highest. These patterns are consistent with the topographical and environmental heterogeneities in the different regions. Western Madagascar has a uniform flat topography, while eastern Madagascar is dominated by the single slope of the NS mountain chain. In contrast, northern Madagascar has a distinctly more heterogeneous environment, including a wide range of different forest types and strong elevation gradients in multiple compass directions. Such heterogeneities likely provide ample opportunity for niche specialization and consequently an elevated speciation rate. Northern Madagascar is a biodiversity hotspot also in many other taxa apart from dung beetles probably for the same reason (Wilme *et al.* 2006, Kremen *et al.* 2008).

To conclude, we have developed a new method for phylogenetic comparative analysis. The use of the state-of-the-art ABC allowed us to implement flexible modelling assumptions. With increasing computation power and availability of phylogenies, questions are becoming more specific and models more complex, but future studies are first required to fully evaluate the strengths and the weaknesses of these inference methods. A central question is to understand which properties of the phylogeny carry the most important information and how adequate is the underlying model in representing true biological processes. As more elaborate models might be needed, and writing the likelihood might be intractable, ABC can be a valuable alternative.

## Acknowledgments

We thank Andreia Miraldo for discussions and Rampal S.

Etienne, Emma E. Goldberg, Daniele Silvestro, Alex Pigot, and one anonymous reviewer for comments on the manuscript. This work was supported by the Academy of Finland (grant numbers 256453 and 250444 to IH).

## References

- Baumont, M. A. 2010: Approximate Bayesian computation in evolution and ecology. — *Annual Review of Ecology, Evolution, and Systematics* 41: 379–406.
- Bloom, D. D., Weir, J. T., Piller, K. R. & Lovejoy, N. R. 2013: Do freshwater fishes diversify faster than marine fishes? A test using state-dependent diversification analyses and molecular phylogenetics of new world silversides (atherinopsidae). — *Evolution* 67: 2040–2057.
- Blum, M. G. B., Nunes, M. A., Prangle, D. & Sisson, S. A. 2013: A comparative review of dimension reduction methods in approximate Bayesian computation. — *Statistical Science* 28: 189–208.
- Buerki, S., Manning, J. C. & Forest, F. 2013: Spatio-temporal history of the disjunct family Tecophilaeaceae: a tale involving the colonization of three Mediterranean-type ecosystems. — *Annals of Botany* 111: 361–373.
- Buerki, S., Jose, S., Yadav, S. R., Goldblatt, P., Manning, J. C. & Forest, F. 2012: Contrasting biogeographic and diversification patterns in two Mediterranean-type ecosystems. — *PLoS ONE* 7(6): e39377, doi:10.1371/journal.pone.0039377.
- Colless, D. 1982: Review of phylogenetics: the theory and practice of phylogenetic systematics. — *Systematic Zoology* 31: 100–104.
- Davis, M. P., Midford, P. E. & Maddison, W. 2013: Exploring power and parameter estimation of the BiSSE method for analyzing species diversification. — *BMC Evolutionary Biology* 13, doi:10.1186/1471-2148-13-38.
- Day, J. 2008: Tempo and mode of diversification of lake Tanganyika cichlid fishes. — *PLoS ONE* 3(3): e1730, doi:10.1371/journal.pone.0001730.
- de Wit, M. J. 2003: Madagascar: Heads it's a continent, tails it's an island. — *Annual Review of Earth and Planetary Sciences* 31: 213–248.
- Etienne, R. S. & Haegeman, B. 2012: A conceptual and statistical framework for adaptive radiations with a key role for diversity dependence. — *The American Naturalist* 180: E75–E89, doi:10.1086/667574.
- Etienne, R. S., Haegeman, B., Stadler, T., Aze, T., Pearson, P. N., Purvis, A. & Phillimore, A. B. 2012: Diversity-dependence brings molecular phylogenies closer to agreement with the fossil record. — *Proceedings of the Royal Society B* 279: 1300–1309.
- FitzJohn, R. G., Maddison, W. P. & Otto, S. P. 2009: Estimating trait-dependent speciation and extinction rates from incompletely resolved phylogenies. — *Systematic Biology* 58: 595–611.
- Freed, L. A., Conant, S. & Fleischer, R. C. 1987: Evolutionary ecology and radiation of hawaiian passerine birds. — *Trends in Ecology & Evolution* 2: 196–203.
- Futuyma, D. J. 1998: *Evolutionary biology*. — Sinauer Associates, Incorporated, Sunderland, MA.
- Goldberg, E. E., Lancaster, L. T. & Ree, R. H. 2011: Phylogenetic inference of reciprocal effects between geographic range evolution and diversification. — *Systematic Biology* 60: 451–465.
- Goodman, S. M. & Benstead, J. P. 2003: *The natural history of Madagascar*. — The University of Chicago Press, Chicago.
- Grant, P. R. 1999: *Ecology and evolution of Darwin's finches*. — Princeton University Press, Princeton, NJ.
- Hanski, I., Wirta, H., Nyman, T. & Rahagalala, P. 2008: Resource shifts in Malagasy dung beetles: contrasting processes revealed by dissimilar spatial genetic patterns. — *Ecology Letters* 11: 1208–1215.
- Jansson, R., Rodriguez-Castaneda, G. & Harding, L. E. 2013: What can multiple phylogenies say about the latitudinal diversity gradient? A new look at the tropical conservatism, out of the tropics, and diversification rate hypotheses. — *Evolution* 67: 1741–1755.
- Kremen, C., Cameron, A., Moilanen, A., Phillips, S. J., Thomas, C. D., Beentje, H., Dransfield, J., Fisher, B. L., Glaw, F., Good, T. C., Harper, G. J., Hijmans, R. J., Lees, D. C., Louis, E., Nussbaum, R. A., Raxworthy, C. J., Razafimpahanana, A., Schatz, G. E., Vences, M., Vieites, D. R., Wright, P. C. & Zjhra, M. L. 2008: Aligning conservation priorities across taxa in Madagascar with high-resolution planning tools. — *Science* 320: 222–226.
- Losos, J. B. & Glor, R. E. 2003: Phylogenetic comparative methods and the geography of speciation. — *Trends in Ecology & Evolution* 18: 220–227.
- Lovette, I. J., Bermingham, E. & Ricklefs, R. E. 2002: Clade-specific morphological diversification and adaptive radiation in Hawaiian songbirds. — *Proceedings of the Royal Society B* 269: 37–42.
- Maddison, W. P. & FitzJohn, R. G. 2015: The unsolved challenge to phylogenetic correlation tests for categorical characters. — *Systematic Biology* 64: 127–136.
- Maddison, W. P., Midford, P. E. & Otto, S. P. 2007: Estimating a binary character's effect on speciation and extinction. — *Systematic Biology* 56: 701–710.
- Mahler, D. L., Ingram, T., Revell, L. J. & Losos, J. B. 2013: Exceptional convergence on the macroevolutionary landscape in island lizard radiations. — *Science* 341: 292–295.
- Miraldo, A. & Hanski, I. A. 2014: Competitive release leads to range expansion and rampant speciation in Malagasy dung beetles. — *Systematic Biology* 63: 480–492.
- Miraldo, A., Wirta, H. & Hanski, I. 2011: Origin and diversification of dung beetles in Madagascar. — *Insects* 2: 112–127.
- Montreuil, O., Viljanen, H. & Miraldo, A. 2014: Evolution of the Malagasy endemic genus *Nanos* Westwood, 1842 (Coleoptera, Scarabaeidae, Epilissini). — *Systematic Entomology* 39: 442–459.
- Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B. & Kent, J. 2000: Biodiversity hotspots for conservation priorities. — *Nature* 403: 853–858.
- Nee, S., May, R. M. & Harvey, P. H. 1994a: The reconstructed evolutionary process. — *Philosophical Transactions of the Royal Society of London B* 344: 305–311.
- Nee, S., Holmes, E. C., May, R. M. & Harvey, P. H. 1994b:

Extinction rates can be estimated from molecular phylogenies. — *Philosophical Transactions of the Royal Society of London B* 344: 77–82.

Numminen, E., Cheng, L., Gyllenberg, M. & Corander, J. 2013: Estimating the transmission dynamics of *Streptococcus pneumoniae* from strain prevalence data. — *Biometrics* 69: 748–757.

Pennell, M. W., FitzJohn, R. G., Cornwell, W. K. & Harmon, L. J. 2015: Model adequacy and the macroevolution of angiosperm functional traits. — *The American Naturalist* 186: E33–E50, doi:10.1086/682022.

Pybus, O. G. & Harvey, P. H. 2000: Testing macro-evolutionary models using incomplete molecular phylogenies. — *Proceedings of the Royal Society of London B* 267: 2267–2272.

Rabosky, D. L. & Lovette, I. J. 2008: Density-dependent diversification in North American wood warblers. — *Proceedings of the Royal Society B* 275: 2363–2371.

Rabosky, D. L. & Goldberg, E. E. 2015: Model inadequacy and mistaken inferences of trait-dependent speciation. — *Systematic Biology* 64: 340–355.

Robert, C. & Casella, G. 2004: *Monte Carlo statistical methods*. — Springer, Science + Business New York.

Rolland, J., Condamine, F. L., Jiguet, F. & Morlon, H. 2014: Faster speciation and reduced extinction in the tropics contribute to the mammalian latitudinal diversity gradient. — *PLoS Biology* 12(1): e1001775, doi:10.1371/journal.pbio.1001775.

Seehausen, O. 2006: African cichlid fish: a model system in adaptive radiation research. — *Proceedings of the Royal Society B* 273: 1987–1998.

Stadler, T. 2013: Recovering speciation and extinction dynamics based on phylogenies. — *J. Evol. Biol.* 26: 1203–1219.

Sunnaker, M., Busetto, A. G., Numminen, E., Corander, J., Foll, M. & Dessimoz, C. 2013: Approximate Bayesian computation. — *PLoS Comput. Biol.* 9(1): e1002803, doi:10.1371/journal.pcbi.1002803.

Toni, T., Welch, D., Strelkowa, N., Ipsen, A. & Stumpf, M. P. H. 2009: Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. — *Journal of the Royal Society Interface* 6: 187–202.

Vences, M., Wollenberg, K. C., Vieites, D. R. & Lees, D. C. 2009: Madagascar as a model region of species diversification. — *Trends in Ecology & Evolution* 24: 456–465.

Verheyen, E. 2003: Origin of the superstock of cichlid fishes from Lake Victoria, East Africa. — *Science* 300: 325–329.

Viljanen, H., Escobar, F. & Hanski, I. 2010: Low local but high beta diversity of tropical forest dung beetles in Madagascar. — *Global Ecology and Biogeography* 19: 886–894.

Waters, J. M., Fraser, C. I. & Hewitt, G. M. 2013: Founder takes all: density-dependent processes structure biodiversity. — *Trends in Ecology & Evolution* 28: 78–85.

Wilme, L., Goodman, S. M. & Ganzhorn, J. U. 2006: Biogeographic evolution of Madagascar’s microendemic biota. — *Science* 312: 1063–1065.

Wirta, H., Viljanen, H., Orsini, L., Montreuil, O. & Hanski, I. 2010: Three parallel radiations of *Canthonini* dung beetles in Madagascar. — *Molecular Phylogenetics and Evolution* 57: 710–727.

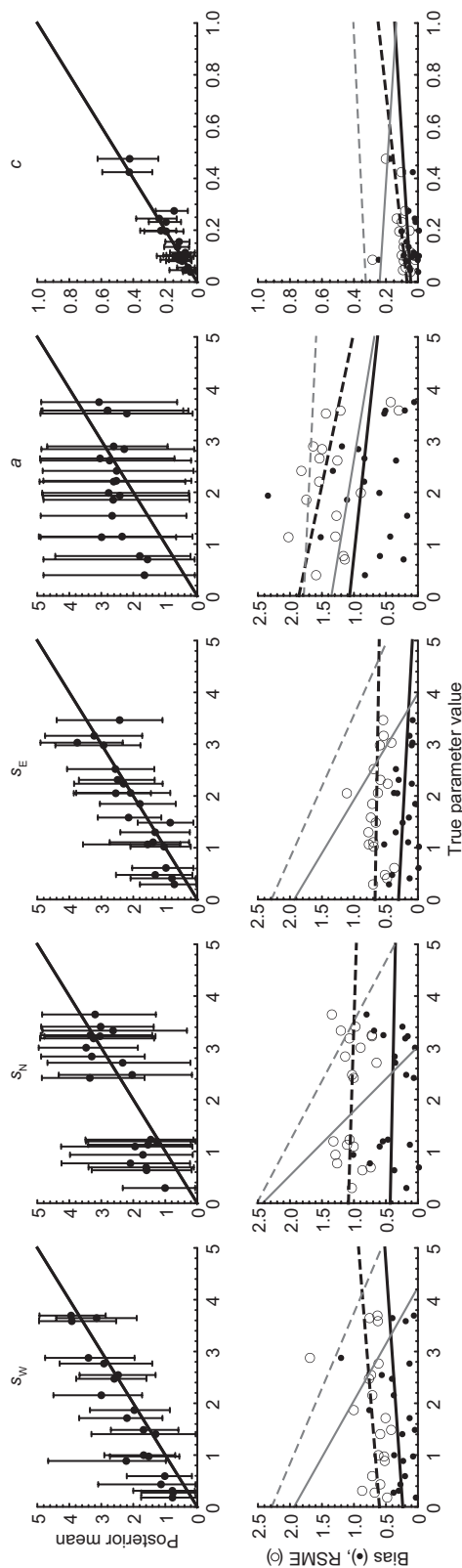
Yoder, A. D. & Nowak, M. D. 2006: Has vicariance or dispersal been the predominant biogeographic force in Madagascar? Only time will tell. — *Annual Review of Ecology Evolution and Systematics* 37: 405–431.

Yule, G. U. 1924: A mathematical theory of evolution based on the conclusions of Dr. J. C. Willis. — *Philosophical Transactions of the Royal Society B* 213: 21–87.

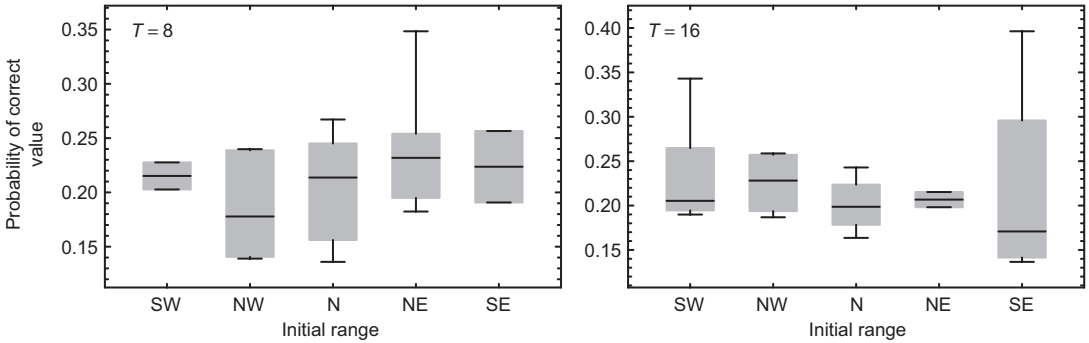
Appendix 1. Inference validation.

**Table A1\_1.** Average Pearson’s correlations (upper triangle) and standard deviations of the correlations (lower triangle) for all the parameters. The values were calculated from the 10 vectors in Fig. 2 in the main text.

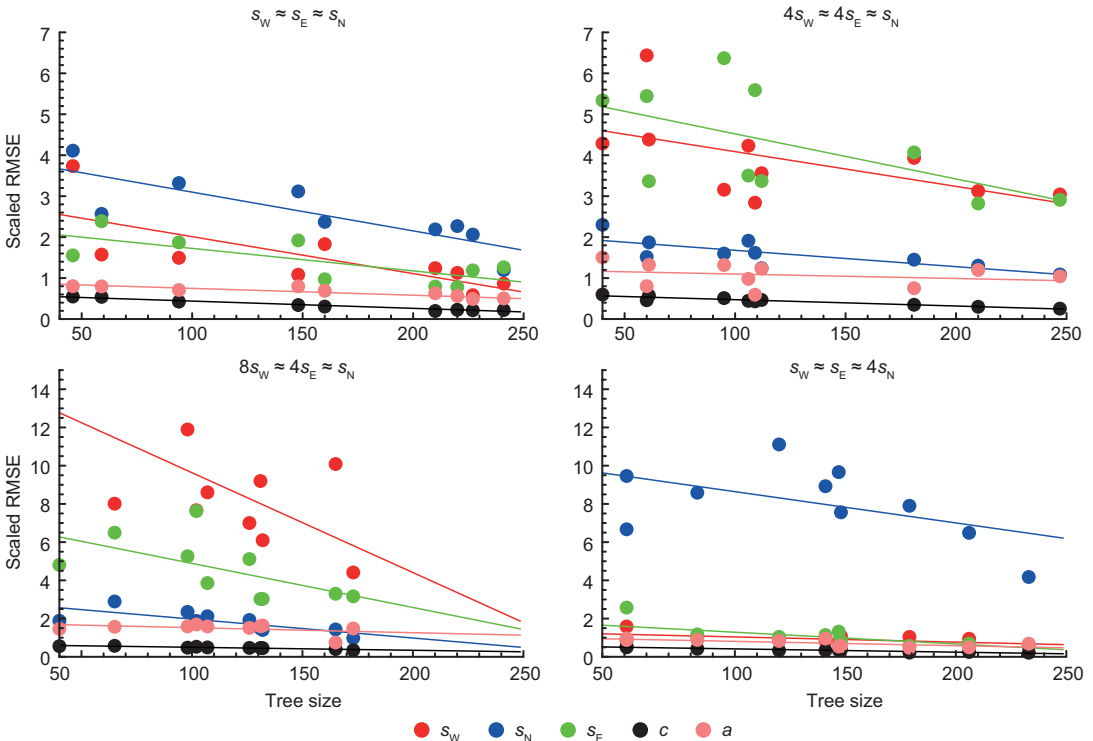
	<i>S<sub>W</sub></i>	<i>S<sub>N</sub></i>	<i>S<sub>E</sub></i>	<i>e<sub>W</sub></i>	<i>e<sub>N</sub></i>	<i>e<sub>E</sub></i>	<i>c</i>	<i>a</i>
<i>S<sub>W</sub></i>	–	–0.097	–0.131	0.458	0.017	–0.032	–0.001	–0.113
<i>S<sub>N</sub></i>	0.057	–	–0.100	–0.044	0.284	–0.029	–0.096	–0.056
<i>S<sub>E</sub></i>	0.080	0.100	–	0.024	–0.005	0.456	–0.035	–0.060
<i>e<sub>W</sub></i>	0.107	0.083	0.096	–	–0.159	–0.124	0.259	0.043
<i>e<sub>N</sub></i>	0.099	0.148	0.092	0.154	–	–0.152	0.221	0.042
<i>e<sub>E</sub></i>	0.059	0.107	0.080	0.099	0.147	–	0.262	0.073
<i>c</i>	0.147	0.087	0.109	0.148	0.139	0.141	–	0.104
<i>a</i>	0.091	0.055	0.116	0.091	0.088	0.107	0.091	–



**Fig. A1\_1.** Inference validation in the simplified scenario for 20 simulated data for  $T = 8$ , in which extinction rate is fixed to 1 for all regions. Columns from left to right represent respectively: within-region speciation rate in the West ( $s_w$ ), the North ( $s_n$ ), and the East ( $s_e$ ), allopatric speciation rate (a), and colonization rate (c). The upper row shows the true value (x-axis) against the mean of the marginal approximate posterior distribution (dot) and its 95% posterior interval (error bar) (c). The line represents the identity line. The bottom row depicts the true parameter value (x-axis) against the root mean squared error (RMSE, circle, y-axis) and the root of the difference between the mean squared error and the posterior variance (bias, dot, y-axis). Black dashed and solid lines are fitted linear regressions of the RMSE and the bias, respectively. For comparison, fitted regressions from Fig. 2 are shown as gray dashed (RMSE) and solid (bias) lines. Simulation parameters:  $N = 400$  and  $q = 50\%$ .



**Fig. A1\_2.** Probability of correctly estimating the initial range of the last common ancestor for  $T = 8$  and 16. The line within the box, the box, and the bars represent respectively the median, the 25%–75% quartiles, and the outer fence.  $N = 400$  and  $q = 50\%$ .



**Fig. A1\_3.** Change in the scaled root mean squared error (RSME) as a function of tree size. Four cases were considered: equal within-region speciation rate in each region; within-region speciation rate highest in North and equal in East and West; within-region speciation rate highest in North and lowest in West; and within-region speciation rate the lowest in North and equal in East and West. In each scenario, 10 samples were drawn keeping the ratio between the within-region speciation rates roughly constant and fixing colonization and between-region speciation rate to constant values.  $T = 8$ ,  $N = 400$ , and  $q = 50\%$ .

Appendix 2. Application to empirical data.

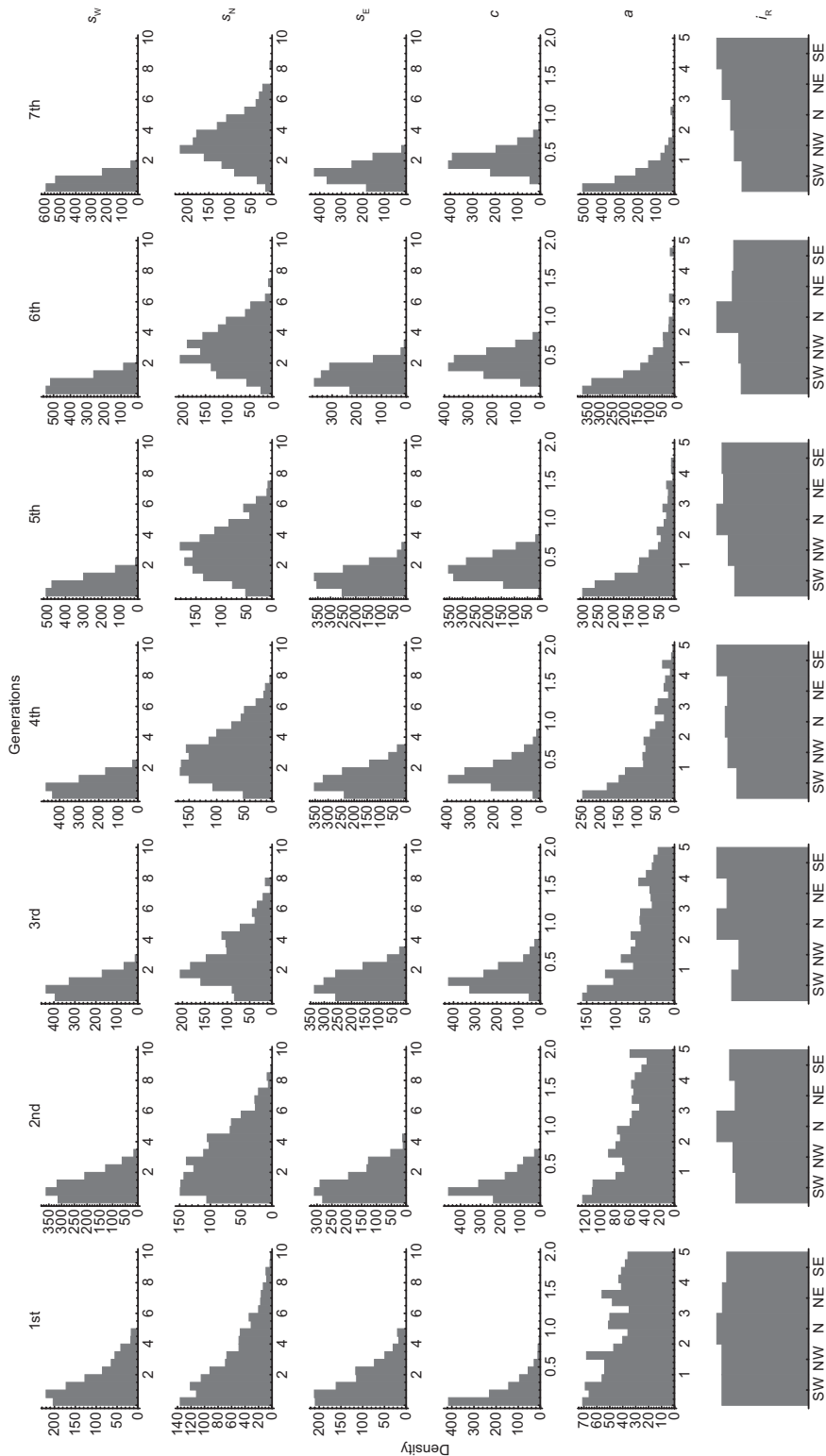
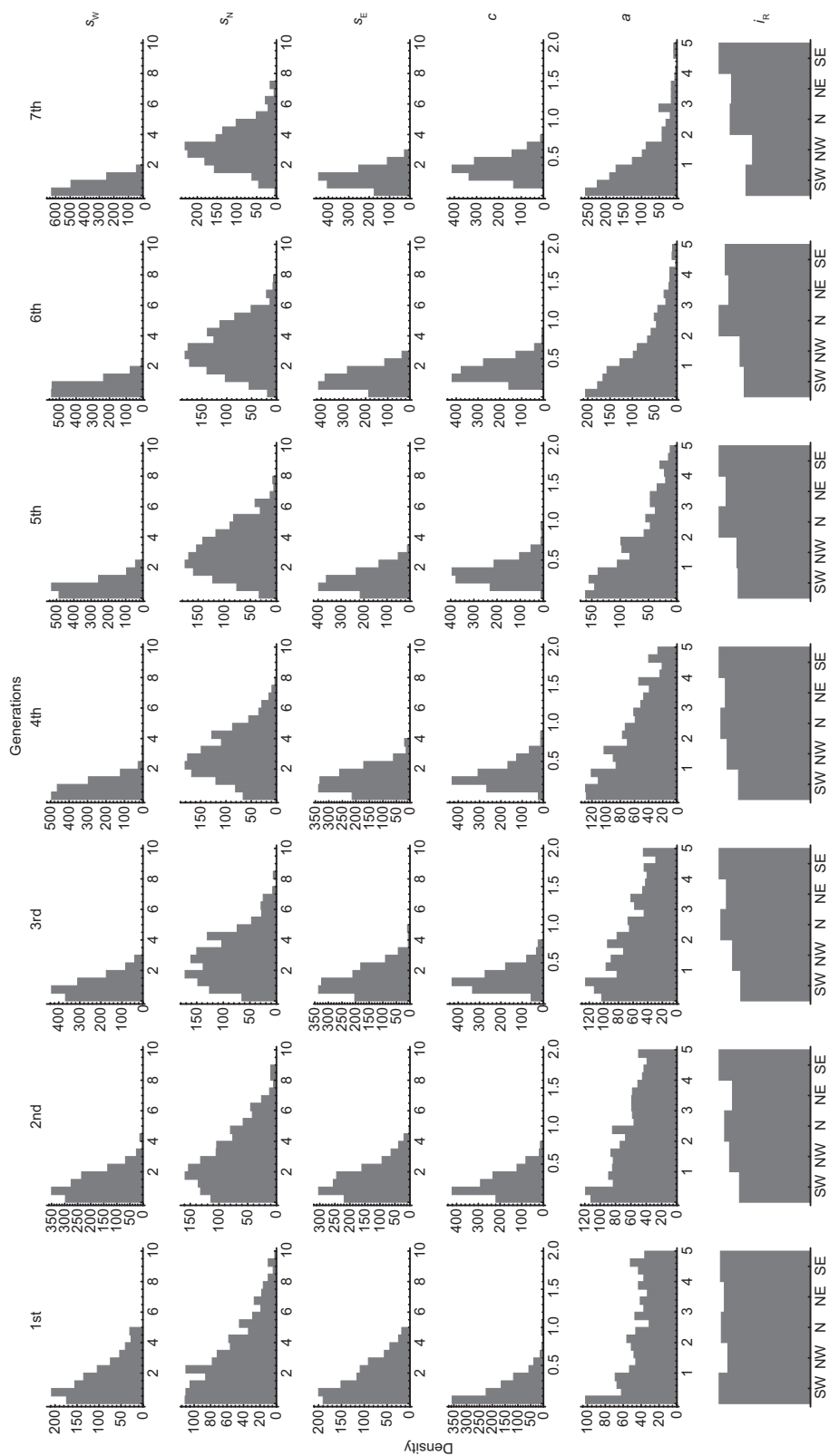


Fig. A2\_1. Convergence of the approximate posterior distributions under the simplified scenario with  $T = 8$ . Other parameters:  $N = 1000$  and  $q = 50\%$ .





**Fig. A2\_2.** Convergence of the approximate posterior distributions under the simplified scenario with  $T = 16$ . Other parameters:  $N = 1000$  and  $q = 50\%$ .

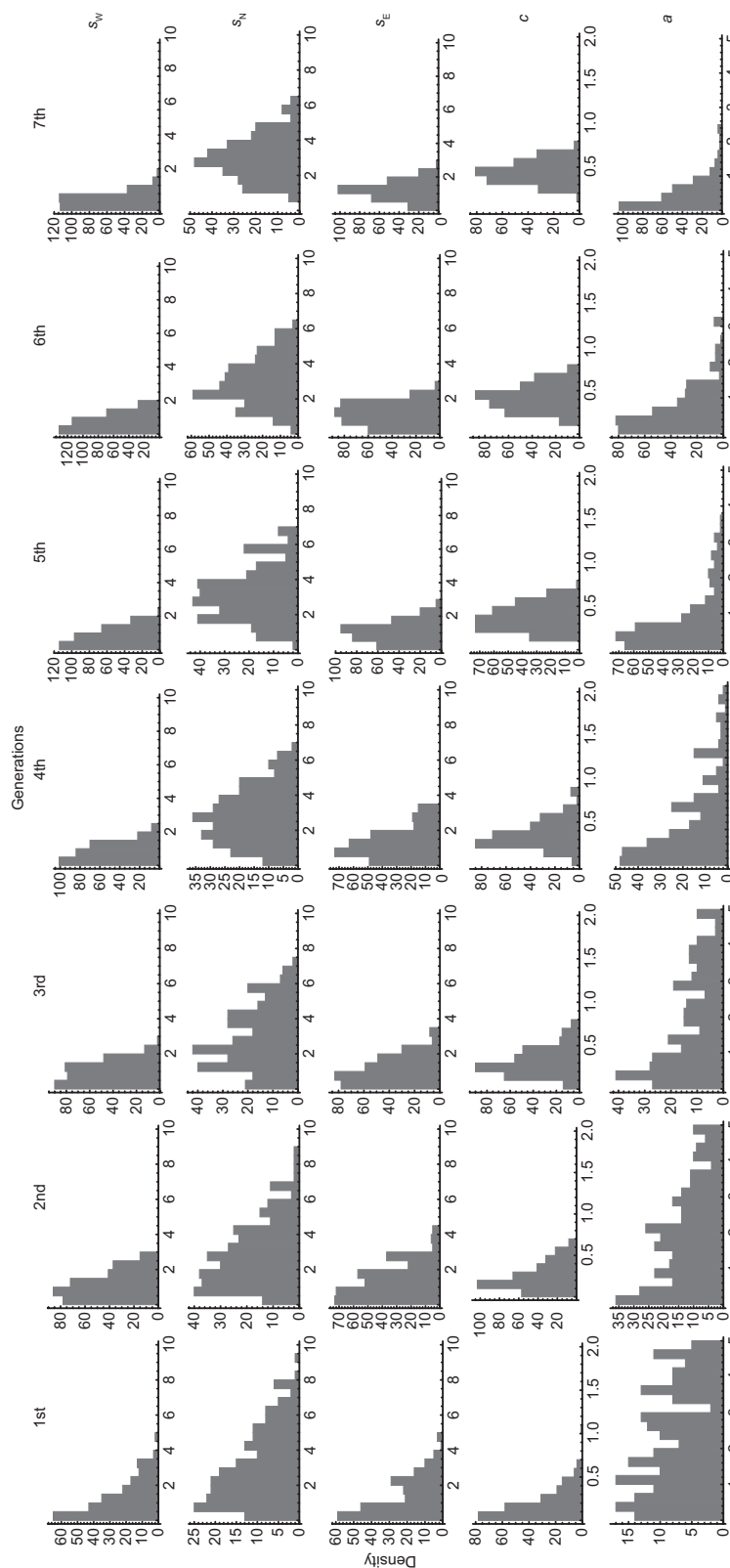


Fig. A2\_3. Convergence of the approximate posterior distributions under the simplified scenario with  $T = 16$  constraining the initial range to North. Other parameters:  $N = 220$  and  $q = 50\%$ .